



From the American College of Epidemiology

## Ethics, big data and computing in epidemiology and public health

Jennifer Salerno PhD<sup>a,\*</sup>, Bartha M. Knoppers PhD<sup>b</sup>, Lisa M. Lee PhD<sup>c</sup>,  
WayWay M. Hlaing PhD<sup>d</sup>, Kenneth W. Goodman PhD<sup>e</sup><sup>a</sup> National Collaborating Centre for Methods and Tools, McMaster University, Hamilton, Ontario, Canada<sup>b</sup> Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, Quebec, Canada<sup>c</sup> Presidential Commission for the Study of Bioethical Issues, Washington DC<sup>d</sup> Division of Epidemiology and Population Sciences, Department of Public Health Sciences, University of Miami, Miami, FL<sup>e</sup> Miller School of Medicine, Institute for Bioethics and Health Policy, University of Miami, Miami, FL

## ARTICLE INFO

## Article history:

Received 2 January 2017

Accepted 4 May 2017

Available online 10 May 2017

## Keywords:

Epidemiology

Ethics

Big data

Public health

Genomics research

## ABSTRACT

**Purpose:** This article reflects on the activities of the Ethics Committee of the American College of Epidemiology (ACE). Members of the Ethics Committee identified an opportunity to elaborate on knowledge gained since the inception of the original Ethics Guidelines published by the ACE Ethics and Standards of Practice Committee in 2000.

**Methods:** The ACE Ethics Committee presented a symposium session at the 2016 Epidemiology Congress of the Americas in Miami on the evolving complexities of ethics and epidemiology as it pertains to “big data.” This article presents a summary and further discussion of that symposium session.

**Results:** Three topic areas were presented: the policy implications of big data and computing, the fallacy of “secondary” data sources, and the duty of citizens to contribute to big data. A balanced perspective is needed that provides safeguards for individuals but also furthers research to improve population health. Our in-depth review offers next steps for teaching of ethics and epidemiology, as well as for epidemiological research, public health practice, and health policy.

**Conclusions:** To address contemporary topics in the area of ethics and epidemiology, the Ethics Committee hosted a symposium session on the timely topic of big data. Technological advancements in clinical medicine and genetic epidemiology research coupled with rapid advancements in data networks, storage, and computation at a lower cost are resulting in the growth of huge data repositories. Big data increases concerns about data integrity; informed consent; protection of individual privacy, confidentiality, and harm; data reidentification; and the reporting of faulty inferences.

© 2017 Elsevier Inc. All rights reserved.

## Introduction

It has been more than 15 years since the original 2000 American College of Epidemiology (ACE) Ethics Guidelines [1] were published. Since then, specialized fields of epidemiology (e.g., genetic and molecular epidemiology) have emerged, and as has awareness that epidemiology is closely interconnected with other fields (e.g., health information technology, global health and noncommunicable diseases). These advances have changed the profession of epidemiology, introducing numerous concepts related to big data and computing. Since the Guidelines' original publication, additional ethical issues in the context of specialized fields of

epidemiology have emerged and presented challenges. To address this need, the Ethics Committee hosted a symposium session at the 2016 Epidemiology Congress of the Americas held in Miami, FL, June 21–24, 2016. This article presents a summary and further discussion of that symposium session. The session addressed three topics: (1) the international policy and human rights implications of big data and computing (B.M.K.); (2) the fallacy of “secondary” data sources (L.M.L.); and (3) the benefits, risks, and duties of citizens to contribute to big data (K.W.G.). This article exemplifies the Ethics Committee's ongoing consultative efforts to highlight contemporary topics in the area of ethics and epidemiology relevant to professional epidemiologists. We are targeting a diverse audience of health care researchers including epidemiologists, health care informatics specialists, geneticists, health care providers, policy-makers, and others who work with, or are interested in working with big data.

\* Corresponding author. National Collaborating Centre for Methods and Tools, McMaster University, McMaster Innovation Park, Suite 210A, 175 Longwood Rd S, Hamilton, Ontario, L8P 0A1. Tel.: (905) 525-9140, ext. 20454; fax: (905) 529-4184.  
E-mail address: [salerno@mcmaster.ca](mailto:salerno@mcmaster.ca) (J. Salerno).

## Big data: potential and challenges

The current technological landscape permits the digitization and storage of unprecedented amount of data from many sources, including smart phones, text messages, credit card purchases, online activity, electronic medical records, and global positioning system data. Many of these data sources contain personal information both related and unrelated to health, including for example, geographic location, health or social security number, and credit card number. Various forms of health information are being easily created, stored, and accessed. The use of genomic information is at the forefront of the big data debate, as it is increasingly recognized that deoxyribonucleic acid [DNA]–based information (e.g., ribonucleic acid [RNA] measurements, single-nucleotide polymorphism [SNP] profiles) can potentially be used to identify individuals [2,3]. Big data provide researchers and private interests the ability to match or link records across a number of data sources. Linking of big data sources of health and heritable information offers great promise for our understanding of disease predictors. It also poses risks to individuals by uncovering negative or potentially discriminatory health-related findings. Wide-scale linkage of big data has implications for privacy, data security, and the procedures for and governance of informed consent.

For the professional practice of specialized fields of epidemiology such as molecular and genetic epidemiology, challenges in the current technological landscape are at the forefront. The challenges are interconnected and include: (1) the unparalleled generation of human genetic data, (2) the enormous health, diagnostic, and therapeutic potential of using human genetic data, and (3) the ability to share these data globally. Taken together, these issues challenge epidemiologists with reconciling the potential benefits of science and the seamless mapping of disease risk with safeguards needed to uphold the long-held ethical principles at the root of the professional practice of epidemiology, including an individual's privacy, confidentiality, and autonomy.

For epidemiologists, use of big data is typically referred to as “the exploration and interpretation of very large and complex data sets derived from pooling of cohorts, the omics technologies, electronically stored medical records, and numerous emerging vehicles for social media and communication” [4]. We propose the idea of “bigger data” to represent the rapidly growing addition of sources of data beyond what was initially recognized, including financial information, health care encounters, pharmaceutical prescriptions, geospatial data, and shopping habits, among others. We propose as the main distinction between big data and bigger data the idea that sources of data and even big data are not static entities. New sources of data, with differing comprehensiveness are likely to be continuously generated and accumulated. The concept of big data has been revolutionary in recent years, and an evolution that has pushed epidemiologists to think beyond traditional sources of data.

## Evolving epidemiology data sources

Historically, we have divided the sources of health information into two categories: primary and secondary. “Primary data” refer to data collected for a specific research question using an instrument (e.g., a survey or laboratory test) designed or chosen to optimize validity. “Secondary data” refers to existing data that were collected for a purpose other than the specific research question at hand. Secondary data might come from routine public health surveillance, population-based health surveys, hospital discharge data, health care administration and billing, or even other research projects. Often an epidemiologist turns to existing data sources to answer research questions more efficiently and to reduce costs. In

the era of big data, long-term retention of both biological samples (e.g., blood or tissue) and related molecular, genetic, demographic, and health outcome data has resulted in the creation of biobanks and data repositories. These information-rich data are available for purchase and use thus allow researchers to bypass the time-consuming and expensive steps of creating data collection instruments as well as recruiting subjects, performing/optimizing laboratory tests, and the longitudinal tracking for outcome information.

The evolution of the technological landscape has given rise to the growth of linkable data sources and direct sources of health data. Increasing storage and use of biospecimens as a rich source of health data suggests that for the professional practice of epidemiology, the distinction between routinely conceived primary and secondary data sources is diminished. REB/IRB members should understand the risks and benefits of conducting epidemiological research in the era of big data and recognize that the source of data can no longer be used as a measure of potential harm to research participants. Epidemiologists and research ethics board (REB)/institutional review board (IRB) members have a professional obligation to understand the ethical dimensions of the potential informational benefits and harm that big data present. We must prepare epidemiologists with the ethical tools they need to work in the new big data era. Explicit ethics training is no longer optional and should be part of every epidemiologist's training. In addition, REBs/IRBs should familiarize themselves with big data and the ethical dimensions of its use.

A critical ethical concern is the potential impact on autonomy when data are linked across multiple data sources without the individual or consumer's permission or informed consent. The potential harm for the individual is the real risk of uncovering one's health information as it relates to personal, socioeconomic, or other determinants of health. In the absence of consistent privacy and confidentiality regulations or rules, and without security safeguards for data sharing opportunities, the potential harm to the individual is in tension with the potential benefits that might result from the research. This risk is increased when the data linkage occurs in the absence of a relevant research question or without a clear benefit to public health. The potential harms or benefits to individuals as a result of the generation of large and linkable data repositories might be unclear or go unnoticed during a REB or IRB review if board members are unfamiliar with the nuances of this new research avenue.

## Evolving access and regulatory landscape

The motivation for the use of big data includes the efficiencies gleaned from creating “economies of scale.” First, data are rapidly generated from genetic, medical, socioeconomic, social media, and geospatial sources; disease and other types of registries; primary care and community clinics; and from data sources that include air pollution, climate, and contaminated soils and water. Second, these data are able to be stored in internet-based centers which would allow government agencies the potential to fund the storage of large data sets in the most rigorously safeguarded and/or the most widely accepted virtual center(s) thereby diminishing overall costs to single researchers and institutions. Third, the aforementioned centralized model of internet-based virtual “cloud” services offer (1) the potential to provide a harmonized and common point of access for large amounts of stored data, (2) greater efficiency in terms of cost and time since each researcher potentially could avoid individualized transfer of large amounts of data between repositories and a model of decentralized cloud services, (3) greater computing efficiencies, by allowing researchers to use as many computers as needed to complete an analysis with only paying for

the computing time accrued, and (4) the ability to produce otherwise unattainable scientific findings [5]. The ease with which numerous data sources can be accessed, downloaded, and analyzed is the catalyst for the cloud's increased utility in the life sciences and medicine, including precision medicine. Epidemiologists aim to use the cloud to analyze data to identify new or stronger associations for benefits to public health.

One important challenge is how to protect data shared on a global scale while maintaining the opportunity to discover and refine scientific findings that lead to health benefits. To maximize a cloud commons approach, a number of challenges remain, including (1) the risk to privacy for individuals who provide data, (2) the potential for broad linkage that jeopardizes social rights such as health care, welfare, housing, employment, education, and equal treatment, (3) the loss of data control, including data integrity and data availability issues, and (4) variability in the ability to safeguard owing to broad government practices. A number of commercial cloud service providers exist and have considered the necessary safeguards such as encrypted data, firewall systems, and control practices for authorization, access and monitoring of data use. However, there are no overarching standards of practice for cloud service use among researchers and institutions [5]. In this endeavor, epidemiologists and researchers must discuss and address the above listed risks anew, as existing privacy and other data-use regulations were drafted long before big data and the cloud were a reality.

Big data and the cloud, and how their emergence might affect our research and practice, have motivated some epidemiologists to advocate for increased clarity of the legal and regulatory landscape. Traditionally, epidemiologists have reserved their role as advocates to confronting public health problems and supporting members of affected communities [1]. Today, one challenge is shaped by the tension between promoting data protection for individuals and maximizing access to this rich data source for health researchers. The tension is increased by the fact that data privacy and other regulations have not kept pace with consumer interests, big data, and the cloud. How should we address the tension between and among the genomic cloud, biomedical research community, individual privacy, and data security?

One proposed approach is to examine current regulations and existing models to identify and catalogue successful practices. Currently, there is the pending new U.S. rule on the "Federal Policy for the Protection of Human Subjects" which was embarked on to revise and updates the longstanding "Common Rule." The final rule further ensures that human subjects' research oversight remains clear to investigators, facilitates valuable research, and promotes seamless policies across U.S. departments and agencies [6]. To the extent that the final rule has implications for the big data landscape, the final rule allows for broad consent practices. Specifically, the final rule allows prospective research to take place using participant data without the researcher's need to seek further informed consent from those participants who provided broad consent for the unspecified further use of their data. This includes subsequent research using and linking identifiable private information from participants and identifiable biospecimens. Consistent with the U.S. notion of a broad consent policy, the European General Data Protection Regulation outlines ways that biomedical research can move forward with the use of genetic and health data while simultaneously upholding safeguards for an individual's protection and privacy (e.g., broad consent, coding/encryption safeguards, data storage conditions, and the ability in certain circumstances to reprocess data without re-contacting subjects) [7]. Moreover, in 2016, the European Safe Harbour policy was replaced with the "E.U.-U.S. Privacy Shield," a framework intended to enhance safeguards in transatlantic data

transfers. Recognition of another country's privacy protections is based on the concept of "adequate levels of protection" as a central mechanism of approval for trans-border data sharing (<https://www.privacyshield.gov/welcome>). Other initiatives of interest include the Commission High Level Expert Group on the European Open Science Cloud, which produced its first draft in June of 2016 to "seamlessly integrate existing networks" and provide a unified "virtual environment to store, share, and re-use data across disciplines and borders" (<http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>). The Canadian-led Cancer Genome Collaboratory (<http://www.genomecanada.ca/en/cancer-genome-collaboratory>) has set-up a cloud facility to enable research on a comprehensive cancer genome set while safeguarding against personal re-identification and addressing other privacy and security issues. The U.S. National Cancer Institute's (NCI) Genomic Data Commons is providing a repository based on NCI-generated data to enable data sharing among researchers in the area of precision medicine and oncology (<https://gdc.cancer.gov/about-gdc>). Addressing privacy and security of cloud computing requires attention to interjurisdictional policies, identifiability of participants including robust but not burdensome access restrictions, secure remote computing techniques, and economies of scale that improve security while reducing burdens on centers and researchers.

### Ethical principles revisited

The ethical dimensions of big data and population health research are not unlike the common ethical principles in epidemiology research and practice. Whatever our data source, we must uphold the ethical principles that reflect what we value—minimizing harms while maximizing benefits, ensuring just distribution of burdens and benefits, respect individual autonomy through informed consent, privacy and confidentiality, build trust, and maintain scientific rigor [1]. To honor these ethical commitments in the big data era, we must be aware of new ways that stored data can cause harm as well as ways they can confer benefit.

To better address the challenges of big data, possible next steps include renegotiating with the public the terms of data collection, storage, and use. For example, researchers and other end-users of data should work with communities to create new norms that include an expectation on the part of communities that they will contribute to a learning health system [8] and a reciprocal expectation on the part of researchers and data users that data will be protected from unauthorized disclosure. Protecting data from disclosure will require that we raise professional expectations of all parties who "touch" data—from researchers to data analysts to information technology professionals—and that we create comprehensive privacy and access policies that include swift and severe consequences for misuse or disclosure of data.

### Societal contributions to big data

Epidemiology has always been information-intensive. Indeed, its very existence depends on the collection and analysis of data and information. Much of that data and information relates, pertains, or is somehow linked to individual people, their families, or their communities. Given the goals and successes of epidemiology and other population health sciences and the sustainability and quality of health care systems, one should infer that the collection and analysis of data and information is itself a moral obligation. Put differently: the failure to collect and analyze data and information with the goal of improving population health would be blameworthy [9].

Epidemiologists collect very large amounts of data and information and, indeed, should be seen as key sources of the vast repositories of big data. Fortunately, many if not most of the ethical issues related to big data are familiar to epidemiologists: privacy and confidentiality, the risks and management of subgroup stigma, the scope of informed consent, the challenge of high-stakes decision-making under uncertainty, and so on. Of special interest for the ethics-and-epidemiology community is the question of whether or when these issues represent something novel and thus require additional ethical analysis. It might, for instance, be the case that big data take us beyond our way of thinking about data that it reshape the ethical issues related to the use of information. That in turn has a significant bearing on public policy.

Epidemiology is successful—just ask the sources of its data and information: our communities, individuals, and special groups of people. These community actors trust the scientists who collect, analyze, and offer advice based on it. This, in turn, entails that those involved in what is known as “public health informatics” [10] must be committed to ensuring data security, preventing public disclosure of personal information and, generally, practicing responsible stewardship. Likewise, individuals in a health care system have a reciprocal duty not to oppose, and perhaps even to facilitate the credible collection and analysis of their data and information for their own treatment and that of others who follow them. When they fail to respect this duty as citizens benefiting from health care knowledge, or when an REB/IRB suggests that privacy and consent rules require the permission of individuals for such collective analysis—they become a barrier to the traditions, systems, and processes that have already improved their individual health and well-being.

### Ethical framework to address big data

Traditional approaches to the issues of big data have relied on ethical principles requiring protection from presumed harm from biomedical research. A new and more positive approach to address the challenges of big data emphasizes human rights. Indeed, this is the basis for the Framework for Responsible Sharing of Genomic and Health-Related Data established by the Global Alliance for Genomics and Health (GA4GH; <https://genomicsandhealth.org/>). At the core of the Framework is the understanding that data sharing of genomic and other health-related data is fundamental to advancing human health and well-being. The Framework is guided by internationally recognized human rights perspectives as they pertain to privacy, nondiscrimination, fair access, and procedural fairness, all based on respect for human dignity. The human rights perspective for data sharing is set out in the Framework, such that it balances the rights of data producers and users to be recognized for their contributions to research while at the same time supporting the rights of those who contribute their data. In this Framework, data producers and users are obligated to engage in responsible scientific conduct and data access, and share data appropriately.

The Framework takes as its starting point the right of all citizens to benefit from advances in science and its applications, or to freely share in scientific advancement and its benefits; and the right to recognition such that every author has the right to protection of the moral and material interests resulting from any scientific, literary, or artistic production (Universal Declaration of Human Rights (1948), Articles 27 [1], and 27 [2]). Grounding international data sharing in a human rights perspective provides or identifies a universalizing force, political and legal dimensions in addition to the moral appeals of bioethics, international legal accountability, group and individual rights, and positive duties and obligations of governments and private actors [11].

Ahead of its time, the 2000 ACE Ethics Guidelines supported such a human rights perspective. The Guidelines outline the ethical duties and obligations and professional responsibilities of epidemiologists including, “to protect the welfare and rights of research participants and to help ensure that the potential benefits of epidemiologic research and practice are maximized and distributed in an equitable fashion.” In the Guidelines, virtuous traits of character are identified for professional epidemiologists in their research and practice such as “humility, fidelity, justice, patience, industry, and veracity.” In addition, the Guidelines emphasize the professional expectation that epidemiologists assign the correct attribution of scientific contributions [1].

The traditional notion of an epidemiologist’s obligation to protect research participants from harm should not be the only ethical dimension re-interpreted and applied to the use of big data. Protection from harm can be balanced against the human right to benefit from scientific advances, and for scientists and data curators to receive proper attribution for their work. Adding a human rights perspective to the big data debate offers a positive framework by which actionable solutions in the form of relevant policies can be developed (e.g., consent policy, accountability policy, privacy and security policy, data sharing lexicon).

### Conclusion

The purpose of this article is a “call to action” in the area of ethics, big data, and computing in epidemiology and public health. Also, to provide readers with information about the activities of the college and to give readers of the *Annals of Epidemiology* a broad perspective on a recent major epidemiological issue. Our intentions for this article and subsequent ethics and epidemiology publications as part of the work of the Ethics Committee are for these resources to be nimble, accessible, and augment the foundations of ethical practice for professional epidemiologists that were established in the original Guidelines. The vision for this article and for prior related published work by the Ethics Committee [12] is for these materials to be used as a resource companion or supplemental information to the original Ethics Guidelines when challenges in the area of ethics and epidemiology are encountered in professional practice. With emerging opportunities for future research using big data (including routine health, administrative, and social data), epidemiologists should be educated on the ethical challenges associated with use of these data. To this end, the ACE Ethics Committee has synthesized and made freely available online more than 30 course syllabi from U.S. academic institutions at the graduate level pertaining to the teaching of ethics and epidemiology and/or public health (<http://www.miami.edu/index.php/ethics/projects/EPHES/>). Other North American ethics teaching resources are provided by the Resources for Research Ethics Education (<http://research-ethics.net/>) and the Canadian Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans [13] which includes corresponding online ethics learning opportunities and ethics capacity building (e.g., learning modules, webinars, and workshops) (<http://www.pre.ethics.gc.ca/eng/education/tutorial-didacticiel/>). The National Collaborating Centers for Public Health also provide resources [14] as well as ethics teaching and training for public health practice (<http://www.ncchpp.ca/55/Ethics.ccnpps>). As such, we recommend that the concepts and teaching resources presented in this article be used by professional epidemiologists and others including trainees in epidemiology in their daily activities or as teaching tools in government, academia, industry, and public health practice. The expanded role of teaching in ethics and epidemiology as it pertains to big data is also an opportunity for



researchers who have experience in using big data and its challenges to mentor future researchers wishing to use such data so that potential for high impact epidemiological research can be realized.

### Next steps

Our plenary session dealt with ethics, big data, and computing in epidemiology and public health from the perspective of our speakers including research and teaching expertise in the areas of ethics, epidemiology, genomics, health policy, legal dimensions, bioethics, and bioinformatics. Additional perspectives may not have been adequately addressed in the plenary session including but not limited to the ethical considerations surrounding study methods/design, data collection/analysis, standards of practice, and policy setting and implementation in the era of big data [15,16]. The discussion presented in this article highlights and embodies the wide reaching and longstanding ethical principles, including those outlined in the 2000 ACE Ethics Guidelines, of data integrity; informed consent; protection of individual privacy, confidentiality, and harm; data re-identification; and the reporting of faulty inferences. This article also presented a big data framework based on a human rights perspective. While the exact mechanisms, rules of engagement, and standards of practice for the use of big data and cloud services are still developing, the new policies and initiatives highlighted in this article by the United States, Canada, and Europe clearly advances and modernizes our ability to conduct high quality and technologically advanced research to improve human health while at the same time upholding the long held ethics principles outlined in this article and in the ACE 2000 ethics guidelines. The future of ethics, big data, and computing in epidemiology and public health holds great promise to improve health and requires a thoughtful approach to keeping that promise.

### References

- [1] American College of Epidemiology. American College of Epidemiology Ethics Guidelines, Available at: [https://acepidemiology.org/ACE/PolicyStatements/Ethics\\_Guidelines.aspx](https://acepidemiology.org/ACE/PolicyStatements/Ethics_Guidelines.aspx). [accessed 21.02.17].
- [2] Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science* 2013;339:321–4.
- [3] Schadt EE. The changing privacy landscape in the era of big data. *Mol Syst Biol* 2012;8:612.
- [4] Hiatt RA, Sulsky S, Aldrich MC, Kreiger N, Rothenberg R. Promoting innovation and creativity in epidemiology for the 21st century. *Ann Epidemiol* 2013;23:452–4.
- [5] Stein LD, Knoppers BM, Campbell P, Getz G, Korbel JO. Data analysis: create a cloud commons. *Nature* 2015;523:149–51.
- [6] Federal policy for the protection of human subjects. 2017. Available at: <https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf> [accessed 21.02.17].
- [7] Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. *Lancet* 2016;387:1374–5.
- [8] Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. *Ann Intern Med* 2012;157:207–10.
- [9] Goodman KW. Ethics, medicine, and information technology: Intelligent machines and the transformation of health care. Cambridge: Cambridge University Press; 2016.
- [10] Goodman KW, Meslin EM. Ethics, information technology and public health: duties and challenges in computational epidemiology. In: Magnuson JA, Fu PC, editors. *Public health informatics and information systems*. 2nd ed. London: Springer-Verlag; 2014. p. 191–209.
- [11] Knoppers BM, Harris JR, Budin-Ljøsne I, Dove ES. A human rights approach to an international code of conduct for genomic and clinical data sharing. *Hum Genet* 2014;133:895–903.
- [12] Salerno J, Hlaing WM, Weiser T, Striley C, Schwartz L, Angulo FJ, et al. Emergency response in a global health crisis: epidemiology, ethics, and ebola application. *Ann Epidemiol* 2016;26:234–7.
- [13] Canadian Institutes of Health Research. Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. 2014. Available at: [http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS\\_2\\_FINAL\\_Web.pdf](http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf).
- [14] National Collaborating Centre for Methods and Tools. Framework for managing ethical challenges in public health practice and program planning. Hamilton, ON: McMaster University; 2017 (Updated 01 February, 2017). Available at: <http://www.nccmt.ca/resources/search/269> [accessed 23.05.17].
- [15] Khoury MJ. Planning for the future of epidemiology in the era of big data and precision medicine. *Am J Epidemiol* 2015;182(12):977–9.
- [16] Khoury MJ, Ioannidis JP. Big data meets public health. *Science* 2014;346(6213):1054–5.